A 3D Steganalytic Algorithm and Steganalysis-Resistant Watermarking

Ying Yang, Ruggero Pintus, Holly Rushmeier and Ioannis Ivrissimtzis

Abstract—We propose a simple yet efficient steganalytic algorithm for watermarks embedded by two state-of-the-art 3D watermarking algorithms by Cho et al. The main observation is that while in a clean model the means/variances of Cho et al.'s normalized histogram bins are expected to follow a Gaussian distribution, in a marked model their distribution will be bimodal. The proposed algorithm estimates the number of bins through an exhaustive search and then the presence of a watermark is decided by a tailor made normality test or a t-test. We also propose a modification of Cho et al.'s watermarking algorithms with the watermark embedded by changing the histogram of the radial coordinates of the vertices. Rather than targeting a continuous statistics such as the mean or variance of the values in a bin, the proposed watermarking modifies a discrete statistic, which here is the height of the histogram bin, to achieve watermark embedding. Experimental results demonstrate that the modified algorithm offers not only better resistance against the steganalytic attack we developed, but also an improved robustness/capacity trade-off.

Index Terms—Polygonal meshes, data embedding, watermarking, steganalysis.

1 Introduction

Digital watermarking is the process of embedding digital signals into digital media such as images, audio, video, or 3D models. In a relationship analogous to that between cryptography and cryptanalysis, as a counterpart to watermarking, *steganalysis* aims at the detection of watermarks hidden into digital signals.

Thus far, numerous excellent watermarking techniques for inserting watermarks into 3D models have been proposed [1], [2], [3], [4], [5]. For a more exhaustive comparison of 3D watermarking methods, we refer the reader to the survey in [6]. These algorithms are primarily concerned with the robustness of the watermark, targeting applications such as proof of ownership and copy control, while the undetectability of the embedded watermark does not seem to be a major concern in their evaluation. That also means that 3D model steganalysis is an underdeveloped area and, to the best of our knowledge, there are only a few works [7], [8] in the literature proposing steganalytic algorithms specifically targeting 3D model watermarking. In contrast, the domain of 2D image watermarking has been greatly influenced over the years by the various proposed steganalytic methods [9], [10], [11], [12], [13].

The thesis motivating this paper is that 3D steganalysis can have a similar positive influence on the development of the field of 3D watermarking in at least two ways. Firstly, by opening new applications domains where the understanding of the anti-steganalytic properties of the watermark are either absolutely essential, e.g., covert communication, or at least extremely important, e.g., proof of ownership and copy control through invisible watermarking. Secondly, by motivating the development of improved watermarking algorithms, which are quite likely to have superior properties not only regarding the undetectability of the watermark but also its robustness.

1.1 Overview

Steganalytic approaches are classified into two categories: *specific* and *universal*. The former detects the presence of a message embedded by particular watermarking algorithms, while the latter aims at message detection regardless of the embedding algorithms used. In this paper we propose a specific steganalytic algorithm for determining the presence of a watermark hidden by Cho et al.'s mean and variance based algorithms [14], which are major 3D watermarking techniques with considerable impact on subsequent research [4], [5]. The proposed algorithm exploits the alteration of the model's natural statistics caused by Cho et al.'s watermark insertion method. More specifically, watermarking with Cho et al.'s method makes the distribution of the means and variances of the normalized histogram bins bimodal, while it is expected to be Gaussian before watermarking.

We also propose a blind 3D watermarking algorithm with improved undetectability and robustness performance over Cho et al.'s. The new algorithm embeds the watermark into the histogram of the radial coordinates of the mesh vertices, as Cho et al.'s does. The main difference is that instead of embedding each watermark bit inside a continuous statistics of the model, e.g., the mean or the variance of a normalized histogram bin, we embed it inside a discrete statistic, that is, the difference in the number of elements of two adjacent bins. Experimental results show that the proposed discrete statistic offers improved performance against both the proposed steganalytic attack and standard watermark removal attacks.

Y. Yang and H. Rushmeier are with Department of Computer Science, Yale University, USA. R. Pintusis is with CRS4, Italy. I. Ivrissimtzis is with School of Engineering and Computing Sciences, Durham University, UK.
This work was partially supported by Sardinian Regional Authorities under the VIGEC project.
Manuscript accepted in Jan, 2016.

The main contributions of the paper are:

- A steganalytic algorithm specifically designed against the two watermarking algorithms proposed by Cho et al. [14], which outperforms the *universal* steganalysis proposed in [8].
- A new 3D watermarking algorithm which is more robust than two state-of-the-art techniques [7], [14] in terms of undetectability against the proposed steganalytic attack, offers a better distortion/capacity trade-off and it is robust against standard watermark removal attacks such as smoothing, noise insertion, subdivision and simplification.

These two contributions introduce to the field of 3D watermarking a new approach into the development of new algorithms, which is analogous to the standard paradigm in image watermarking. In that paradigm, watermarking algorithms are developed through a competition between steganographers and steganalysts.

The main limitation of the proposed steganalytic method is its narrow application domain, that is, it is tailored to work against Cho et al.'s watermarking algorithms. Nevertheless, we reasonably expect that the method can also be used against any other method embedding watermarks by changing the means or variances of specific attributes of the model.

1.2 Related Work

Similarly to image watermarking, 3D watermarking techniques are broadly classified into two categories: *spatial-based* and *transform-based*. In the spatial domain, Yeo et al. [15] propose a fragile watermarking method which perturbs a vertex ensuring that predefined hash functions have the same value on it. One of its drawbacks is the causality problem, due to its heavy dependence on the order of traversal of vertices. Lin et al. [2] address this issue using vertex-order-independent hash functions. To increase robustness, Yu et al. [16] and Cho et al. [14], instead of inserting the watermark into a single vertex, embed each watermark bit into a group of vertices. Bors [17] uses a neighborhood localized measure to select the vertices that give small embedding distortion and watermark these vertices by local perturbations. Aiming at robustness against mesh editing or pose deformation, Yang et al. [4] propose a Laplacian coordinates based algorithm where the watermark bits are hidden by altering the histogram of the lengths of the Laplacian vectors.

Frequency analysis based algorithms can achieve excellent results on both watermark robustness and imperceptibility. By using the spectral analysis by Karni et al. [18], Ohbuchi et al. [19] propose an algorithm embedding the watermark into the low frequencies of Karni et al.'s decomposition. This method is non-blind, thus requiring the availability of the original model during watermark extraction. Praun et al. [1] propose a robust, non-blind watermarking method using an edge collapse based multiresolution decomposition. Kanai et al. [20] propose a non-blind method for semi-regular meshes based on the modification of wavelet coefficients, while Uccheddu et al. [21] extend this approach to be a blind one.

Closely related to watermarking are steganographic methods. While the distinction between watermarking and steganography is not sharp, the term steganography is largely used when we favor large embedding capacity, usually at the expense of robustness, while watermarking schemes are evaluated principally on their robustness. Cayre et al. [22] introduce a steganographic algorithm treating each triangle as a two-state object, depending on the position of the projection of a vertex onto its opposite edge. The maximum capacity this method can achieve is 1 bit per vertex. Inspired by this idea, Wang et al. [23] increase the embedding capacity while minimizing the embedding distortion via using a multi-level hiding procedure. Subsequently, they [24] extend their prior work [23], taking into account texture information during embedding. Two steganographic methods include Chao et al. [25] and Yang et al. [26]. All the above mentioned steganographic approaches can achieve high capacity and low distortion, but the main limitation is their weak robustness, that is, they cannot withstand malicious attacks aimed at destroying the embedded message.

The area of steganalysis has been primarily developed on images. Fridrich et al. [27] and Ker [28] propose methods specific for the detection of LSB replacement. Farid [29] proposes a universal approach which uses a wavelet-like decomposition to build higher-order statistical models of natural images. Other universal steganalytic approaches for images include Xuan et al. [30], Wang et al. [12] and Lie et al. [10].

Compared to image steganalysis, 3D steganalysis is a more challenging task since the typical 3D models are more complex objects, having arbitrary topology and irregularly sampled geometry. As such, 3D steganalytic techniques are seriously underdeveloped and considerable research effort is required before approaching a level of maturity similar to that of image steganalysis by solving the various open problems in the field. Farid's method for image steganalysis [29] has been extended in [8] to 3D meshes. In Yang et al. [7], a specific 3D steganalysis is proposed based on the observation that the natural statistics of the 3D model are disturbed by Cho et al.'s mean based embedding. The current paper is an extended version of [7] and includes a new watermarking scheme with smaller distortion compared to [7], [14] and better steganalytic properties that make it practically undetectable by the developed steganalytic attack. The embedding algorithm is based on histogram shape modification, similarly to some existing image watermarking algorithms [31], [32].

2 STEGANALYTIC ALGORITHM

In this section, we first briefly describe the two watermarking algorithms by Cho et al. [14] and then present in detail the proposed steganalytic attack against these two methods.

2.1 Cho et al.'s Watermarking Algorithms

Both mean-based and variance-based variants of Cho et al.'s method embed the message by perturbations of the radial coordinates of the vertices in a spherical coordinate system.

Mean-based watermarking: First, the radial coordinates of all the vertices are computed with respect to the barycenter

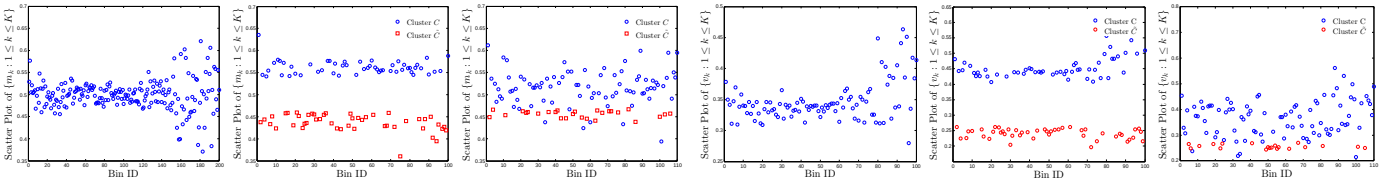


Fig. 1. Scatter plot of the mean values $\mathcal{M} = \{m_k : 1 \leq k \leq K\}$ (three left figures) and the variance values $\{v_k : 1 \leq k \leq K\}$ (three right figures) for the clean *Bunny* with $K = 100$ bins, and the marked *Bunny* with correct estimate $K = 100$ bins and with incorrect estimate $K = 110$ bins.

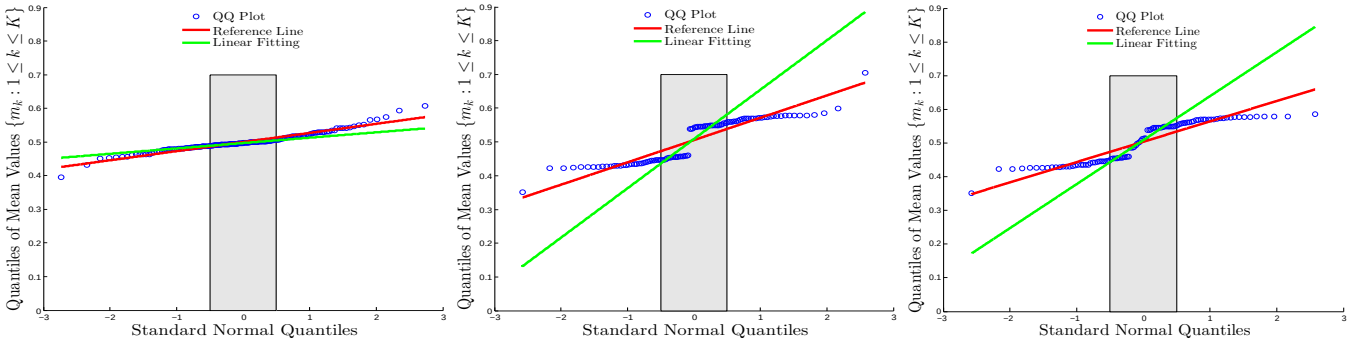


Fig. 2. Q-Q plot of the mean values $\mathcal{M} = \{m_k : 1 \leq k \leq K\}$ for the clean *Rabbit* (left), the marked by Cho et al.'s method with 100 bins (middle) and the marked by the proposed discrete variant of the method with 100 bits (right).

of the set of vertices of the 3D mesh model. Then, a K bin histogram of the radial coordinates is constructed and the elements in each histogram bin are normalized in the interval [0, 1]. Let $\hat{\mathcal{B}}_k = \{\hat{\rho}_{k,j} : j = 1, 2, 3, ...\}$ denote the k-th $(1 \leq k \leq K)$ bin of the normalized radial coordinates $\hat{\rho}_{k,j}$. They embed a -1 (+1) watermark bit in that bin by perturbing the mesh vertices to make the mean value m_k

$$m_k = \frac{1}{|\hat{\mathcal{B}}_k|} \sum_j \hat{\rho}_{k,j} \tag{1}$$

smaller (respectively, greater) than 1/2. Here, $|\cdot|$ stands for the number of elements in a set.

Variance-based watermarking: Similarly, the variance-based version of the algorithm starts with building a K bin histogram of the radial coordinates, but the elements $\hat{\rho}_{k,j}$ in each bin this time are normalized into [-1, 1] rather than [0,1]. Finally, a -1 (+1) watermark bit is embedded into each normalized bin $\hat{\mathcal{B}}_k$ by perturbing the mesh vertices to make the variance

$$v_k = \frac{1}{|\hat{\mathcal{B}}_k|} \sum_j \hat{\rho}_{k,j}^2 \tag{2}$$

smaller (respectively, greater) than 1/3, assuming that the mean of the elements $\hat{\rho}_{k,j}$ is zero.

2.2 The Proposed Steganalytic Algorithm

Our steganalytic algorithm is based on the observation that the mean-based and variance-based embeddings of Cho et al.'s algorithm result in a 2-clustering of the set of the mean values $\mathcal{M} = \{m_k : 1 \leq k \leq K\}$ or variances $\mathcal{V} = \{v_k : 1 \leq k \leq K\}$, respectively, (see Fig. 1). For brevity, we will denote either of these two sets as $\mathcal{M}/\mathcal{V}$ when the distinction is not necessary.

Estimation of *K*: The main challenge towards a fully automatic steganalytic attack against Cho et al.'s methods is finding the number K of histogram bins that was used in the phase of watermark embedding. If a wrong value of K is picked for the reconstruction of the histogram bins for the purposes of steganalysis, then the distribution of $\mathcal{M}/\mathcal{V}$ will be indistinguishable between watermarked and clean models. That is, with an incorrect K, the two clusters C and $\tilde{C}$ of $\mathcal{M}/\mathcal{V}$ will not be well separated even for watermarked models (see Fig. 1).

Consequently, the first step of the proposed steganalytic algorithm finds an estimate of K obtained by an exhaustive search through all possible values. Using a standard clustering algorithm based on Expectation Minimization (EM), for each K^* we classify the elements of $\mathcal{M}/\mathcal{V}$ into two clusters C and $\tilde{C}$ by fitting a mixture of two Gaussians $\mathcal{N}(\mu_{K^*,i}, \sigma^2_{K^*,i}), i = 1, 2$ [33]. We estimate the degree of separation between C and $\tilde{C}$ by the Bhattacharyya distance D_{K^*} [34] between the two Gaussians of the mixture model

$$D_{K^*} = \frac{1}{4} \frac{(\mu_{K^*,2} - \mu_{K^*,1})^2}{\sigma^2_{K^*,1} + \sigma^2_{K^*,2}} + \frac{1}{2} \ln \frac{(\sigma^2_{K^*,1} + \sigma^2_{K^*,2})}{2\sigma_{K^*,1}\sigma_{K^*,2}} \tag{3}$$

and estimate K by

$$K = \arg\max_{K^*} \{D_{K^*} : K^* \in [K_{\min}, K_{\max}], K^* \in \mathbb{N}\} \tag{4}$$

subject to

$$\text{abs}\,(|C| - |\tilde{C}|)/K \leq \epsilon. \tag{5}$$

$K_{\min}$ and $K_{\max}$ define the range of K we would like to consider; here we fix $K_{\min} = 30$ and $K_{\max} = 500$. ϵ in Eq. 5 is a user-specified constant preventing the selection of a very uneven clustering with respect to the size of the two clusters. The justification of the constraint is the assumption that the watermark bits follow a uniform random distribution [35], and hence we expect $|C| \approx |\tilde{C}|$. Without the constraint, the distance maximization in Eq. 4 might return as optimal

a clustering consisting of a small cluster containing a few outliers and a large cluster with all the other values.

Given the estimate of K, we employ a tailor made *normality test* or a standard *t-test* to make a steganalysis decision, that is, to decide whether $\mathcal{M}/\mathcal{V}$ are samples from a single Gaussian, in which case we have a clean model, or they come from a mixture of two Gaussians, in which case we have a watermarked model.

Normality Test: Even though standard normality tests exist, here we use a test specifically designed for the extreme cases we deal with. Indeed, since K is selected for making the distribution of $\mathcal{M}/\mathcal{V}$ as bimodal as possible, a less sharp test may reject the normality assumption even in clean meshes.

We make use of the Q-Q plots, which plot the quantiles of two distributions against each other. The first distribution is the empirical distribution of $\mathcal{M}/\mathcal{V}$ and the second is the standardized normal distribution. If two distributions are linearly related, here if $\mathcal{M}/\mathcal{V}$ are linearly related to the normal distribution, the points in the Q-Q plot are nicely modeled by the *reference line* [36]

$$y = \sigma \cdot x + \mu$$

where μ and σ are the mean and the standard deviation of $\mathcal{M}/\mathcal{V}$.

We check if the reference line is a good model of the points of the Q-Q plot by comparing it with the least square linear fit of these points. That is, we compute the angle of these two lines

$$\theta = \mathrm{abs}\left(\arctan(\sigma) - \arctan(s)\right) \tag{6}$$

where $\mathrm{abs}\,(x)$ stands for the absolute value of x; s denotes the slope of the least square linear fit, and compare it against a threshold θ_T. If $\theta > \theta_T$, we assume that the reference line is not a good model of the Q-Q plot, hence $\mathcal{M}/\mathcal{V}$ is not Gaussian and the mesh is marked. Fig. 2 shows the Q-Q plots, the reference line and the least square linear fit for $\mathcal{M}$ of the clean and a marked *Rabbit* model. To reduce the impact of outliers in $\mathcal{M}/\mathcal{V}$, we compute the least square linear fit from the points in the range [-0.5, 0.5] of the normal distribution in the Q-Q plot, the gray area in Fig. 2.

***t*-test:** In contrast to the normality test which assumes that the set $\mathcal{M}/\mathcal{V}$ of a clean model follows a single Gaussian distribution, t-test is based on the hypothesis that the two clusters C and $\tilde{C}$ are independent random samples from Gaussian distributions with equal means. The main observation is that while C and $\tilde{C}$ of unmarked models should have certain overlap, C and $\tilde{C}$ of marked models are expected to be clearly separated. Consequently, the rejection of the hypothesis of equal means would imply a marked model.

Thus, in this second variant of the steganalytic algorithm, the steganalytic decision is based on a t-test with significance level α, with null hypothesis that the data in C and $\tilde{C}$ of $\mathcal{M}/\mathcal{V}$ are independent random samples from normal distributions with equal means and unknown variances, against the alternative that the means are not equal.

3 MODIFIED WATERMARKING ALGORITHM

In a bid to improve the anti-steganalytic properties of Cho et al.'s algorithms [14], we developed a watermarking algorithm utilizing discrete rather than continuous statistics of the histogram of the radial coordinates. Each watermark bit $w_i \in \{-1, +1\}$ is embedded into a pair of bins $(\mathcal{B}_k, \mathcal{B}_{k+1})$ and its value depends on whether the bin $\hat{\mathcal{B}}_k$ is shorter or taller than $\hat{\mathcal{B}}_{k+1}$, that is, on the sign of the difference of the number of elements in the two bins $|\hat{\mathcal{B}}_k| - |\hat{\mathcal{B}}_{k+1}|$. The embedding process manipulates the difference $|\hat{\mathcal{B}}_k| - |\hat{\mathcal{B}}_{k+1}|$ by considering the triplet of neighboring bins $(\mathcal{B}_k, \mathcal{B}_{k+1}, \mathcal{B}_{k+2})$ and possibly transferring among them some elements, carefully selected to minimize embedding distortion.

In a second modification to Cho et al.'s algorithms, the origin O of the spherical coordinate system is not the barycenter of the vertex set, but it is instead a point chosen to improve the trade-off between embedding distortion and robustness.

Compared to the watermarking method proposed in [7], the two main improvements in our current approach are:

- Instead of transferring elements between two bins $(\mathcal{B}_k, \mathcal{B}_{k+1})$, the embedding process transfers elements among three bins $(\mathcal{B}_k, \mathcal{B}_{k+1}, \mathcal{B}_{k+2})$. By considering triplets rather than pairs of bins, we are able to select the to-be-modified elements from a larger set, further reducing the embedding distortion. Notice that as the exact amount of embedding distortion depends on a multitude of factors, such as the the value of K and the specific sequence of watermark bits, the use of three-bin watermarking does not always guarantee lower embedding distortion. However, the experiments have shown that, on average, the three-bin watermarking offers a significant reduction of the distortion as compared to the two-bin method.
- While [7] uses an empirically found point on the principal axis as the origin O of the spherical coordinate system, the current method finds O by solving an optimization problem, resulting in a more evenly distributed histogram of the radial coordinates. Such optimized histogram is able to carry more watermark bits than the one in [7], since it has more embeddable pairs of bins (see Eq. 11). In addition, embedding over more even histograms can reduce the number of element transfers, reducing embedding distortion. Indeed, when $|\mathcal{B}_k| < |\mathcal{B}_{k+1}|$ occurs in a less evenly distributed histogram, we have to transfer more elements in order to achieve $|\hat{\mathcal{B}}_k| > |\hat{\mathcal{B}}_{k+1}|$ and encode a watermark bit.

3.1 Watermark Embedding

The embedding process is described in detail as follows.

Step 1: The first step is to compute the origin of the spherical coordinate system. Let (x_i, y_i, z_i) be the Cartesian coordinates of the i-th vertex of a polygonal mesh with N vertices. We compute the barycenter (x_c, y_c, z_c) of the vertex set and obtain its principal axis $\mathbf{v}$ after applying Principal Component Analysis (PCA) to it. The origin $O = (\bar{x}, \bar{y}, \bar{z})$ of the spherical coordinate system is a point on the principal axis, that is, on line passing through (x_c, y_c, z_c) with the direction $\mathbf{v}$

$$(\bar{x}, \bar{y}, \bar{z}) = (x_c, y_c, z_c) + \bar{t} \cdot \mathbf{v}. \tag{7}$$

$\bar{t}$, which is computed through an exhaustive search, minimizes the total variation in the heights of triplets of neighboring bins

$$\bar{t} = \underset{t\in[t_{\min},t_{\max}]}{\arg\min} \sum_{k=1}^{K-2} \begin{aligned}&\operatorname{abs}(|\mathcal{B}_k^t| - |\mathcal{B}_{k+1}^t|) + \\ &\operatorname{abs}(|\mathcal{B}_k^t| - |\mathcal{B}_{k+2}^t|) + \\ &\operatorname{abs}(|\mathcal{B}_{k+1}^t| - |\mathcal{B}_{k+2}^t|).\end{aligned} \tag{8}$$

$[t_{\min}, t_{\max}]$ defines the range we would consider for t and is computed so that the resulting $(\bar{x}, \bar{y}, \bar{z})$ candidates are within the bounding box of the 3D model; $\operatorname{abs}(x)$ denotes the absolute value of x.

Unlike [7] which uses an empirically selected point as the origin of the spherical coordinate system, the proposed watermarking method chooses for origin a point that gives a histogram with smooth variation of bin frequencies. As mentioned earlier, the two advantages are: (i) more embeddable triplets of bins and (ii) lower embedding distortion.

Following a standard notation [37], the spherical coordinates of the vertex (x_i, y_i, z_i) are denoted by $(\rho_i, \phi_i, \theta_i)$, where $1 \le i \le N$, $\rho_i \in [0, +\infty)$, $\phi_i \in [0, \pi]$ and $\theta_i \in [0, 2\pi)$.

Step 2: We build a histogram with K bins $\mathcal{B} = \{\mathcal{B}_k : 1 \le k \le K\}$ for the radial coordinates $\mathcal{P} = \{\rho_i : 1 \le i \le N\}$. The bin $\mathcal{B}_k$ is the subset of $\mathcal{P}$

$$\mathcal{B}_k = \{\rho_i : \rho_{\min} + (k-1)\cdot\Delta_\rho \le \rho_i < \rho_{\min} + k\cdot\Delta_\rho\} \tag{9}$$

where $\rho_{\min}$ and $\rho_{\max}$ are the minimum and the maximum of $\mathcal{P}$, and

$$\Delta_\rho = (\rho_{\max} - \rho_{\min})/K \tag{10}$$

is the range size of each bin. We also assume that $\rho_{\max} \in \mathcal{B}_K$.

Step 3: Starting from the second bin $\mathcal{B}_2$, we arrange adjacent bins into pairs $(\mathcal{B}_2, \mathcal{B}_3)$, $(\mathcal{B}_4, \mathcal{B}_5)$, $\ldots$, $(\mathcal{B}_{k-2}, \mathcal{B}_{k-1})$ and hide a watermark bit into each *embeddable* pair of bins. A pair is embeddable if

$$|\mathcal{B}_k| + |\mathcal{B}_{k+1}| + |\mathcal{B}_{k+2}| \ge 1 \tag{11}$$

Notice that no watermark bits are carried by the bins $\mathcal{B}_1$ and $\mathcal{B}_K$. That means that both end vertices of the projection on the principal axis do not move during embedding, increasing the robustness under blind extraction. If K is odd, the bin pairing process requires to exclude one more bin; here $\mathcal{B}_{K-1}$.

A watermark bit w_i is embedded in an embeddable pair $(\mathcal{B}_k, \mathcal{B}_{k+1})$ by transferring some elements among the bins of the triplet $(\mathcal{B}_k, \mathcal{B}_{k+1}, \mathcal{B}_{k+2})$ via increasing or decreasing their values. More specifically, to encode/insert $w_i = -1$, we move elements from the bins $\mathcal{B}_k$ and $\mathcal{B}_{k+2}$ into $\mathcal{B}_{k+1}$ until, if possible,

$$|\hat{\mathcal{B}}_{k+1}| - |\hat{\mathcal{B}}_k| \ge n_{\text{thr}}, \tag{12}$$

where $\hat{\mathcal{B}}_k$ is the k-th watermarked histogram bin and $n_{\text{thr}} \ge 1$ is a user specified integer threshold used to control the robustness/distortion trade-off. Specifically, we increase the values of n_{left} elements ρ_i of $\mathcal{B}_k$ according to

$$\rho_i^+ = \rho_{\min}^{k+1} + \frac{\Delta_\rho}{\underset{n\in\mathbb{N},n\ge5}{\arg\min}\{n : \rho_{\min}^{k+1} + \Delta_\rho/n < \rho_{\max}^{k+1}\}}, \tag{13}$$

and decrease the values of n_{right} elements ρ_i of $\mathcal{B}_{k+2}$ according to

$$\rho_i^- = \rho_{\max}^{k+1} - \frac{\Delta_\rho}{\underset{n\in\mathbb{N},n\ge5}{\arg\min}\{n : \rho_{\max}^{k+1} - \Delta_\rho/n > \rho_{\min}^{k+1}\}}, \tag{14}$$

pushing them into $\mathcal{B}_{k+1}$. Here, ρ_i^+ and ρ_i^- are the new radial coordinates and $\rho_{\min}^{k+1}$ and $\rho_{\max}^{k+1}$ are the minimum and the maximum in $\mathcal{B}_{k+1}$, respectively. The denominator of the fraction in Eq. 13 is an integer, chosen such that $\hat{\rho}_i$ is inside the range of the existing elements of $\mathcal{B}_{k+1}$, that is, $\rho_{\min}^{k+1} < \rho_i^+ < \rho_{\max}^{k+1}$, and it is as near to $\rho_{\max}^{k+1}$ as possible. Similarly, Eq. 14 makes ρ_i^- as near to $\rho_{\min}^{k+1}$ as possible.

To minimize the embedding distortion, the n_{left} elements ρ_i of $\mathcal{B}_k$ and n_{right} elements ρ_i of $\mathcal{B}_{k+2}$ to be moved into $\mathcal{B}_{k+1}$ are chosen as those minimizing the distortion measure

$$\sum_{i=1,\rho_i\in\mathcal{B}_k}^{n_{\text{left}}} \operatorname{abs}(\rho_i - \rho_i^+) + \sum_{i=1,\rho_i\in\mathcal{B}_{k+2}}^{n_{\text{right}}} \operatorname{abs}(\rho_i - \rho_i^-). \tag{15}$$

Since $\mathcal{B}_{k+2}$ will be used to carry the next watermark bit w_{i+1} and thus the histogram's even distribution property needs to be maintained, we only consider a subset of $\mathcal{B}_{k+2}$ when moving elements. Here, $\lceil|\mathcal{B}_{k+2}|/8\rceil$ out of $|\mathcal{B}_{k+2}|$ elements can be moved, i.e., $n_{\text{right}} \le \lceil|\mathcal{B}_{k+2}|/8\rceil$.

Finally, we move elements into $\mathcal{B}_{k+1}$ according to the following two cases:

Case 1: If $|\mathcal{B}_{k+1}| - |\mathcal{B}_k| \ge n_{\text{thr}}$, then $n_{\text{left}} = 0$ and $n_{\text{right}} = 0$, meaning no alteration is required.

Case 2: Else we repeatedly transfer an element from $\mathcal{B}_k$ or $\mathcal{B}_{k+2}$ into $\mathcal{B}_{k+1}$, incrementing n_{left} or n_{right} by one, respectively, and we stop when Eq. 12 is satisfied. Notice that this simple incremental process minimizes Eq. 15 and no global optimization over all the elements of $\mathcal{B}_k$ and $\mathcal{B}_{k+2}$ is required.

The embedding process for $w_i = +1$ is very similar and the details are omitted.

Step 4: After embedding the watermark bits into the radial coordinates, we convert the spherical coordinates $(\hat{\rho}_i, \phi_i, \theta_i)$ to the Cartesian coordinates $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ according to Eq. 16

$$\begin{cases} \hat{x}_i = \hat{\rho}_i \cdot \cos\phi_i \cdot \sin\theta_i + \bar{x} \\ \hat{y}_i = \hat{\rho}_i \cdot \sin\phi_i \cdot \sin\theta_i + \bar{y} \\ \hat{z}_i = \hat{\rho}_i \cdot \cos\theta_i + \bar{z} \end{cases}, \tag{16}$$

where $\hat{\rho}_i$ stands for the watermarked ρ_i, producing the watermarked 3D model.

3.2 Watermark Extraction

The watermark extraction process is straightforward and can be performed with no reference to the original 3D model.

Given a marked mesh, we compute the set of watermarked radial coordinates and construct a histogram with K bins $\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_k : 1 \le k \le K\}$. We form the pairs of bins $(\hat{\mathcal{B}}_2, \hat{\mathcal{B}}_3), (\hat{\mathcal{B}}_4, \hat{\mathcal{B}}_5), (\hat{\mathcal{B}}_6, \hat{\mathcal{B}}_7), \cdots$ and count the number elements $|\hat{\mathcal{B}}_k|$ in each bin. Finally, the watermark bits $\hat{w}_i$ are sequentially extracted from each pair $(\hat{\mathcal{B}}_k, \hat{\mathcal{B}}_{k+1})$ by

$$\hat{w}_i = \begin{cases} -1 & \text{if } |\hat{\mathcal{B}}_{k+1}| \ge |\hat{\mathcal{B}}_k| \\ +1 & \text{otherwise} \end{cases}. \tag{17}$$

TABLE 1
Comparison between the universal steganalysis in [8] and the proposed specific steganalysis with $\epsilon = 0.15$ applied on: Cho's et al.'s methods; their variants with $\beta = 10\%$ of the bins left intact; Yang et al.'s watermarking [7]; the proposed watermarking. The fourth column shows the accuracy in the estimation of K and the fifth the accuracy of the final steganalytic decision. In the fifth column, the left and right subcolumns, where applicable, show the accuracy rates corresponding to the tailor made normality test and the t-test, respectively.

Method	#Bits	#Marked 3D	Accuracy of K	Steganalytic accuracy	
[8] (against six watermarking methods)	64	556	N/A	80.32%	
[8] (against mean-based)	64	359	N/A	80.93%	
[8] (against variance-based)	64	353	N/A	92.96%	
Mean-based (original)	64	443	96.84%	98.52%	92.93%
	100	386	96.63%	97.91%	95.94%
Mean-based (variant with $\beta = 10\%$)	64	443	96.84%	98.29%	81.64%
	100	386	96.63%	97.66%	94.96%
Variance-based (original)	64	443	98.65%	99.32%	82.62%
	100	443	93.00%	97.18%	86.57%
Variance-based (variant with $\beta = 10\%$)	64	443	98.65%	99.32%	80.14%
	100	443	93.00%	96.84%	85.33%
[7] (using m_k)	64	439	0.23%	59.93%	49.71%
	100	426	0	57.45%	51.89%
[7] (using $\lvert\mathcal{B}_k\rvert - \lvert\mathcal{B}_{k+1}\rvert$)	64	439	88.00%	70.82%	80.89%
	100	426	93.40%	73.78%	96.06%
Ours (using m_k)	64	443	0.23%	65.99%	51.36%
	100	443	0	61.34%	50.23%
Ours (using $\lvert\mathcal{B}_k\rvert - \lvert\mathcal{B}_{k+1}\rvert$)	64	443	0.68%	60.16%	51.02%
	100	443	0.45%	69.13%	50.23%

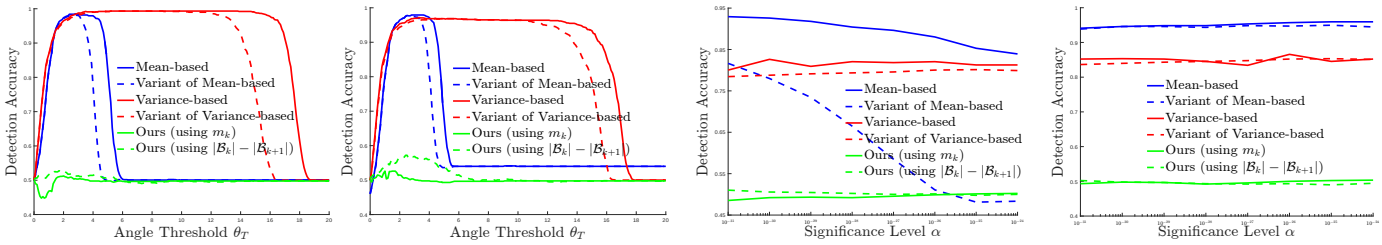


Fig. 3. Plot of the detection accuracy with respect to the angle threshold θ_T (two left figures) and with respect to the significance level α for the t-test (two right figures) for the following methods: Cho et al.'s mean- and variance-based watermarking, their variants and our proposed watermarking, where 64 (left) and 100 (right) bits are embedded into the clean models.

4 EXPERIMENTAL RESULTS

The performance of the proposed steganlytic algorithm against the original Cho et al.'s methods including their variants with some bins left in purpose unmarked and the improved version proposed here is validated in Sections 4.1 and 4.2, respectively. Note that while many steganalytic approaches [27], [29] have been proposed for digital images, to the best of our knowledge, [8] is the only steganalytic methods for 3D models available in the literature. Consequently, we compare the proposed method against the 3D steganalysis in [8] only.

4.1 Steganalytic Algorithm Validation

We validated the steganalytic algorithm on a test set consisting of 445 clean models, mostly from Princeton's University repository [38], and their marked counterparts. As some 3D meshes are unable to carry the watermark for certain values of K, we might have different numbers of marked models for different K's. The results, using a fixed value of $\epsilon = 0.15$ in Eq. 5, are summarized in Table. 1. The sensitivity of the parameter ϵ will be discussed at the end of this subsection, while accuracy rates with ϵ treated as a variable are shown in Fig. 4.

As a first observation, we notice that the proposed method for estimating the number of bins K achieves high accuracy rates for both of Cho et al.'s algorithms. For example, the accuracy rate for the estimation of K reaches as high as 98.65% for the variance based method when embedding 64 watermark bits into the clean models. Due to the successful estimation of K, the final steganalytic decisions are also highly accurate and in particular, as expected, the proposed specific steganalytic methods outperform the universal algorithm [8].

Normality Test: Fig. 3 plots the detection accuracy with respect to the angle threshold θ_T, measured in degrees, for Cho et al.'s methods for watermarks of sizes 64 and 100 bits, respectively. The figure implies that satisfactory steganalysis rate (e.g., $\geq 95\%$) can be obtained when we set the threshold θ_T to any value within the interval [2, 3] for the mean based watermarking and any value within [2, 13] for the variance based watermarking.

By choosing a value of $\theta_T = 2.8$ and $\theta_T = 5.2$, the proposed steganalytic method detects the existence of watermark with accuracy rates of 98.52% for Cho et al.'s mean based algorithm and 99.32% for the variance based algorithm when embedding 64 bits, see Table. 1.

***t*-test:** To justify the use of the tailor made normality for the final steganalytic decision, we compare it against the standard t-test. Fig. 3 plots the detection rates obtained by the t-test versus the significance level α. We notice

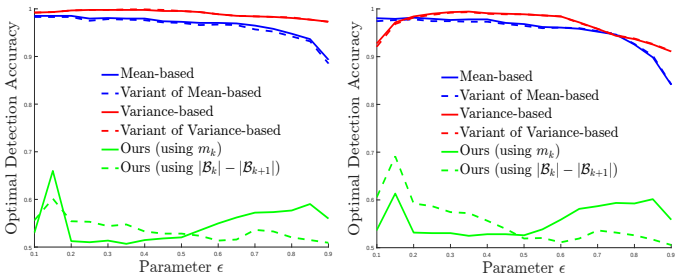


Fig. 4. The accuracy rate plotted against the parameter ϵ, using the tailor made normality test, for 64 (left) and 100 (right) watermark bits.

that when embedding 100 bits we can obtain a detection accuracy above 95%; however, that rate is still inferior the one achieved by the tailor made normality test and, moreover, the highest accuracy rates correspond to extremely small values of the confidence α, raising numerical stability concerns. When embedding 64 bits, the t-test is not able to achieve detection rates above 95% and starts degrading when we use very small values of α.

For a direct comparison of the accuracy rates obtained by the two approaches, in Table 1 we also report the accuracy rates of the t-test for optimal values of α.

Variants of Cho et al.'s: The high detection rates of the proposed steganalytic algorithms are a direct consequence of the clear separation between the clusters C and $\tilde{C}$ (see Fig. 1) in a marked model, when of course the correct value of K is used. Thus, as a quick fix to the problem of improving the steganalytic properties of Cho et al.'s algorithms, one natural idea is to leave a percentage β of the histogram bins intact, i.e., unmarked, to enervate the degree of separation and hence to frustrate the proposed steganalysis. We refer to the algorithms obtained by this technique as the *variants* of the originals. Fig. 2 (right) illustrates the Q-Q plot of the mean values of the *Rabbit* model watermarked by the variant of Cho et al's mean based method with 100 watermark bits and $\beta = 10\%$.

Fig. 3 shows the accuracy rates obtained on the variants of Cho et al.'s with $\beta = 10\%$, using exactly the same steganalytic algorithm we applied on the original. We notice that in conjunction with the tailor made normality test the algorithm is very robust in handling the variants, achieving detection rates very similar to those obtained for the original algorithms. On the other hand, we also notice that the range of values of the threshold θ_T giving near optimal results has been narrowed. The t-test, again, significantly underperformed the tailor made test.

Parameter Sensitivity: The estimation of K involves the use of a parameter ϵ which prevents false choices of K resulting from clusterings consisting of one small cluster of outliers and a much larger cluster with all other values of $\mathcal{M}/\mathcal{V}$. The use of that parameter ϵ is justified by the assumption that the watermark bits are i.i.d. random variables following the Bernoulli distribution with $p = 0.5$, that is, 0.5 probability for either a +1 or a -1 bit, which is generally true in practical applications. To test the robustness of the algorithm against the choice of ϵ we repeated the tests treating ϵ as a variable. The results, obtained with the use of the tailor made normality test, are shown in Fig. 4 and confirm that the proposed steganalysis is largely insensitive to the choice of ϵ.

TABLE 2
Model details and parameter setting.

Model	#Vertices	n_{thr}
Bunny	34835	44
Rabbit	70658	40
Venus	100759	28
Dragon	50000	75
Horse	112642	100

4.2 Watermarking Algorithm Validation

We evaluated the proposed watermarking algorithm on a small representative set of well-known 3D models, consisting of the *Bunny*, *Rabbit*, *Venus*, *Dragon* and the *Horse*, see Fig. 5. The validation tests include *embedding distortion*, *embedding capacity* and *watermark robustness*. The size of the models and the watermarking parameter used are listed in Table 2.

To measure the amount of distortion caused by the embedding of the watermark we utilized two widely used quantitative measures: the root mean square error (RMSE) with respect to the bounding box diagonal and the MSDM2 distance [39] between the original and the marked models. Notice that the MSDM2 distance correlates better with human vision than the Hausdorff distance. The computations were done with the Metro tool [40] and Mepp[1], respectively. Regarding the robustness of the watermark, we used the standard measure of the correlation coefficient

$$\mathcal{C}(\mathbf{w}, \mathbf{w}') = \frac{\sum_i (w_i - \bar{w}) \cdot (w'_i - \bar{w}')}{\sqrt{\sum_i (w_i - \bar{w})^2 \cdot \sum_i (w'_i - \bar{w}')^2}} \qquad (18)$$

where $\bar{w}$ and $\bar{w}'$ are the means of the inserted watermark sequence $\mathbf{w}$ and the extracted sequence $\mathbf{w}'$, respectively.

Embedding Distortion: Fig. 7 plots the RMSE and MSDM2 distances against the number of histogram bins K. We notice that the values of both error measures are low, showing that the proposed watermarking causes only slight distortions to the original carrier models. We also observe that an increase in the value of K generally leads to a decrease of the embedding distortion. The reason is that by increasing K we decrease the bin width Δ_ρ (see Eq. 10) and hence decrease the modulation of the radii ρ_i when they are transferred from one bin to an adjacent.

To gauge the visual significance of the distortion, Fig. 6 shows several marked models for $K = 400$. Any distortion caused by the watermark insertion is hardly noticeable; however, after zooming in we might be able to observe some artifacts in the smooth areas of the model. Fig. 8 shows back to back close-ups of the clean and the watermarked *Bunny* and *Rabbit*.

Embedding Capacity: The theoretical maximum length of a watermark bit sequence that can be embedded is $\lfloor (K-2)/2 \rfloor$. However for large values of K this maximum capacity is usually unachievable because some triplets

1. http://liris.cnrs.fr/mepp/

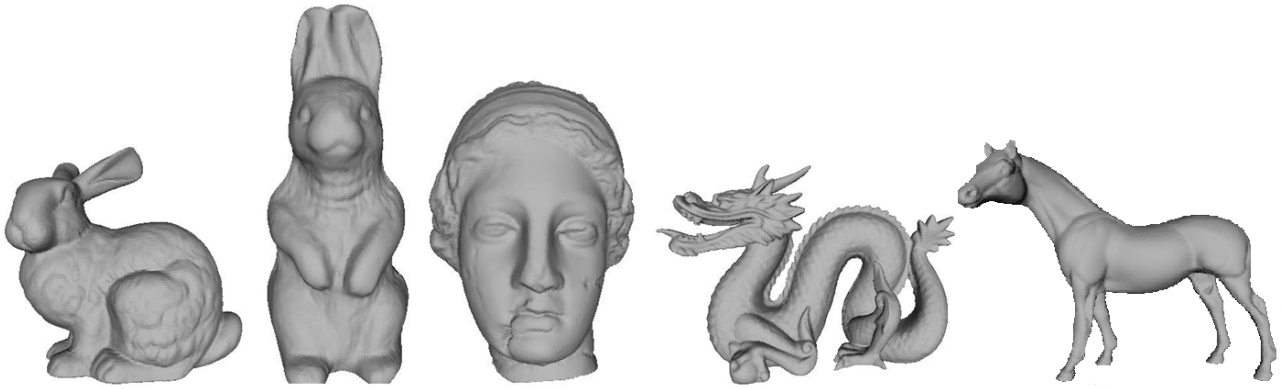


Fig. 5. Original cover mesh models (left to right): *Bunny*, *Rabbit*, *Venus*, *Dragon* and *Horse*.

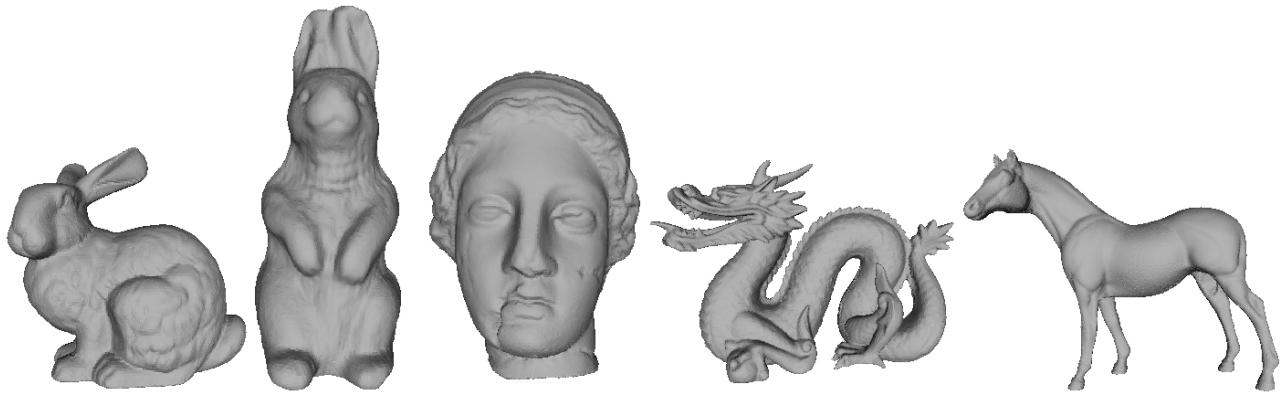


Fig. 6. Watermarked mesh models under $K = 400$ (left to right): *Bunny*, *Rabbit*, *Venus*, *Dragon* and *Horse*.

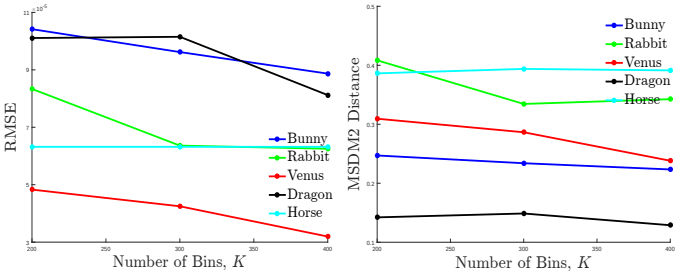


Fig. 7. Embedding distortion measured by RMSE (left) and MSDM2 (right) for $K = 200$, 300 and 400.

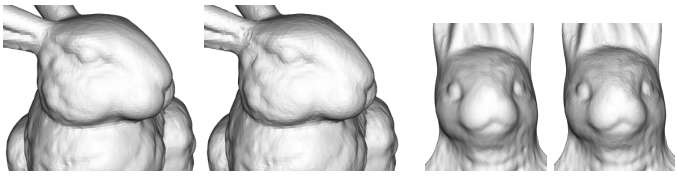


Fig. 8. Close-ups of the original and watermarked *Bunny* and *Rabbit* models. For each mesh group, the left and right figures show the original and marked models, respectively. The watermarked models are those in Fig. 6.

$(\mathcal{B}_k, \mathcal{B}_{k+1}, \mathcal{B}_{k+2})$ do not contain any radii, i.e., $|\mathcal{B}_k| + |\mathcal{B}_{k+1}| + |\mathcal{B}_{k+2}| = 0$, making the pair $(\mathcal{B}_k, \mathcal{B}_{k+1})$ *non-embeddable*.

Watermark Robustness: The proposed method is robust against distortionless operations such as vertex reordering, translation, rotation, uniform scaling and their combinations, due to the invariance of the histogram of the radii ρ_i under such transformations.

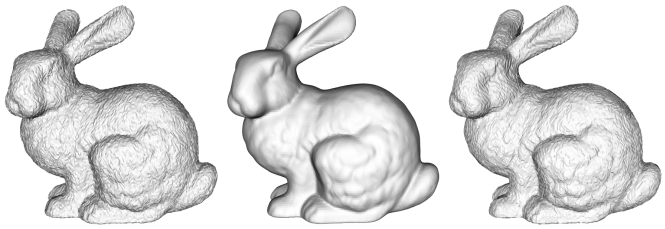


Fig. 9. From left to right: results of attacking the marked *Bunny* model by adding $A = 0.50\%$ noise, carrying out 50 iterations of Laplacian smoothing ($\lambda = 0.02$) and performing 8-bit quantization.

Watermark robustness was also evaluated against common malicious attacks, such as *noise addition*, *smoothing*, *quantization*, *subdivision* and *quadric edge collapse*. We tested against these attacks, fixing $K = 400$ and varying strength of the attack, following the mesh watermarking benchmark [41]. Fig. 9 shows the marked *Bunny* model under various attacks.

Noise Addition: Random noise was added to all vertex coordinates (x_i, y_i, z_i) according to (resp. y_i, z_i)

$$x_i' = x_i + a_i \cdot \bar{d} \tag{19}$$

where $\bar{d}$ is the average radial coordinate, and a_i is a uni-

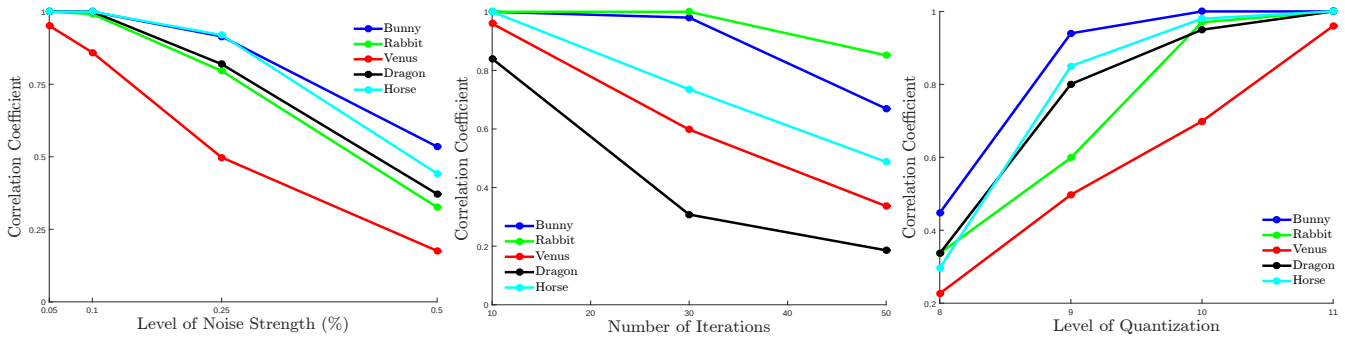


Fig. 10. Robustness against different types of attacks. From left to right: the figures show the correlation coefficients for noise addition, smoothing and quantization.

formly random number in the interval $[-A, A]$. We tested on four different levels of noise : $A = 0.05\%, 0.10\%, 0.25\%$ and 0.50% of the average of the radial coordinates. For each level, we conducted five experiments with different seeds, generating five distinctive noise patterns. Fig. 10 (left) the average correlation coefficient with respect to the level of noise.

We notice that even for a noise level as high as 0.25%, the correlation $\mathcal{C}(\mathbf{w}, \mathbf{w}')$ has a value that is larger than 0.75 for all the test models expect *Venus*. Notice that a noise attack at the level of 0.25% will degrade significantly the visual quality of the carrier model, meaning that even though the watermark may be removed, the model will be of no use to the attacker. We also notice that the *Bunny* and *Rabbit* models appear to be exceptionally robust against the noise attack, retaining a $\mathcal{C}(\mathbf{w}, \mathbf{w}') \approx 0.92$ even when the noise level reaches $A = 0.25\%$. We believe that this is due to the fact that *Dragon* is the least spherical shape and thus, its radii histogram has the highest variance.

Smoothing Attack: To evaluate robustness against *smoothing*, we applied to the marked models 10, 30 and 50 iterations of Laplacian smoothing [42], fixing the deformation factor at $\lambda = 0.02$. The results are shown in Fig. 10 (middle). The correlation coefficients for the *Bunny* and *Rabbit* models are above 0.98 and equal to 1, respectively, even after 30 smoothing iterations. Again, the results imply that the proposed watermarking method is able to survive mesh smoothing as well.

Quantization Attack: We quantized the marked models at 8, 9, 10 and 11 bits and tried to retrieve the embedded watermark from the quantized marked models. Fig. 10 (right) shows the correlation coefficients. As expected, the robustness decreases with the level of quantization, however, the correlation coefficients are satisfactory even for a relatively coarse quantization. In particular, the correlation coefficient of the 11 bit quantization is equal to 1 in four out of the five test models.

Subdivision Attack: We subdivided the watermarked models using the open-source software MeshLab [2]. Table 3 lists the correlation coefficients for five test models. We see that the proposed watermarking is robust against the interpolating Midpoint and Butterfly subdivision schemes, with the correlation coefficient equal to 1 after three subdivision steps, while it is less robust against subdivision with the approximating Loop scheme.

2. http://meshlab.sourceforge.net/

TABLE 3
Correlation coefficients for various watermarked models after subdivision. We used 3 iterations for Midpoint and Butterfly subdivision and 1 iteration for the Loop's LS3 subdivision.

Model	Butterfly	Midpoint	LS3
Bunny	1.0	1.0	0.99
Rabbit	1.0	1.0	0.97
Venus	1.0	1.0	0.72
Dragon	1.0	1.0	0.73
Horse	1.0	1.0	0.97

TABLE 4
Correlation coefficients for various watermarked models simplified at different reduction rates.

Model	10%	20%	30%
Bunny	1.0	1.0	1.0
Rabbit	1.0	1.0	0.98
Venus	1.0	1.0	0.92
Dragon	1.0	1.0	1.0
Horse	1.0	0.95	0.94

Simplification Attack: We evaluated the robustness of the algorithm against simplification with the quadric edge collapse algorithm implemented in Meshlab. As Table 4 shows, the correlation coefficients range from 0.92 to 1, indicating that our proposed watermarking algorithm is quite robust even when 30% of the vertices have been removed.

Watermark Overwriting: Given a marked 3D model, one could attempt to embed a new watermark, overwriting the old one. From the perspective of the owner of the 3D model, the insertion of a new watermark with the same number K of histogram bins and the same origin O would replace the old watermark, erasing it completely. However, being equivalent to running the embedding algorithm twice, watermark overwriting would most probably lead to a model with higher distortion as compared to simply discarding the marked model and inserting the new watermark on the original clean model.

From an attacker's perspective, while the ability to overwrite the watermark would enable unauthorized use of the model, the main technical challenge is that both the bin number K and the origin O are unknown to them and although algorithms for estimating them can be developed,

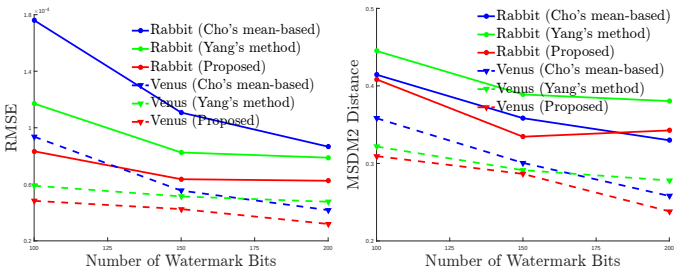


Fig. 11. Experimental comparison between Yang et al.'s method [7], Cho et al.'s method [14] and the proposed method in terms of embedding distortion measured by the RMSE (left) and the MSDM2 distance (right) when embedding 100, 150 and 200 watermark bits.

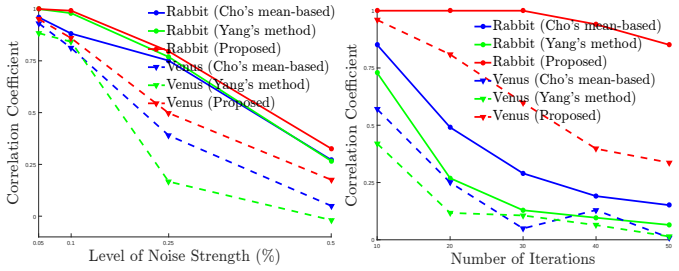


Fig. 12. Experimental comparison between Yang et al.'s method [7], Cho et al.'s method [14] and the proposed method in terms of robustness against noise addition attacks (left) and Laplacian smoothing attacks with $\lambda = 0.02$ (right), when embedding 200 watermark bits.

accuracy could be a problem. For example, Table. 1 shows that the accuracy of the estimates of K obtained by the proposed steganalytic method is unsatisfactory.

4.3 Comparison of Watermarking Algorithms

Finally, we compare the proposed watermarking against the algorithms by Yang et al. [7] and Cho et al. [14] in terms of embedding distortion and robustness against both common malicious attacks and steganalysis. We also compared to [8] for evaluating the anti-steganalysis performance. In all our experiments, for the Yang et al.'s and Cho et al.'s methods we use the parameter settings recommended in [7], [8] and [14].

Embedding Distortion: Fig. 11 (left) shows that our algorithm generally produces lower RMSE and MSDM2 distances than Yang et al.'s [7] and Cho et al.'s [14], implying smaller amounts of distortion.

Watermark Robustness against Malicious Attacks: Fig. 12 shows that, again, our method consistently obtains higher correction coefficients than [7] and [14] when the same noise addition or smoothing attacks are applied, meaning that the watermark bits are recovered with lower error rates.

Anti-steganalysis: When testing how well the proposed watermarking method resists the developed steganalytic attack, we applied the steganalysis not only on the distribution of the means of the bins $\{m_k : 2 \leq k \leq K-1\}$, but also on the more relevant in this case discrete statistic of differences $\{|\mathcal{B}_k| - |\mathcal{B}_{k+1}| : 2 \leq k \leq K-1\}$, trying again to detect a possible bimodality on their distribution.

From Table 1 and Fig. 3, we see that the proposed steganalytic method can break Cho et al.'s original algorithms [14] with accuracy rates of up to 99%, including the cases where some bins are left in purpose without watermark bits. Regarding the watermarking method in [7], the proposed steganalysis generally does not detect watermarks with any satisfactory accuracy, the only exception being when the t-test based normality test is applied to the statistic $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$. However, its behavior is unstable as the steganalytic accuracy varies from 80.89% for 64 bit watermarks to 96.06% for 100 bit watermarks.

Regarding the watermarking method proposed here, with accuracy rates in the region 50%-60%, the proposed steganalytic method fails to detect a watermark, no matter which of the two tests (tailor made normality test, or t-test) is used, and no matter which statistic (the means of the bins m_k, or the differences of the heights of neighboring bins $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$) is targeted. Moreover, Fig. 4 confirms that this failure of the steganalytic attack cannot be fixed by suitable choices of ϵ.

The above results show that the main objective of the development of the new watermarking algorithm has been achieved. Indeed, compared to the original Cho et al.'s algorithms, the proposed watermarking method exhibits significantly improved steganalytic properties, at least as far as resistance against the developed steganalytic attacks is concerned.

Discrete vs continuous statistics: In [14], the explicitly stated underlying assumption for the distribution of the radial coordinates is that the elements of each normalized bin follow the uniform distribution $U(0,1)$. Under this assumption, which is not a requirement for the working of the algorithm but it is essential for low distortion, the means of the normalized bins follow the Bates distribution

$$\frac{1}{|\hat{\mathcal{B}}_k|}\sum_{i=1}^{|\hat{\mathcal{B}}_k|} U(0,1).$$

On the other hand, under the same assumption, in a pair of adjacent bins of equal width and with $n = |\hat{\mathcal{B}}_k| + |\hat{\mathcal{B}}_{k+1}|$ elements in total, the number of elements in the first bin follows the binomial distribution $B(n, 1/2)$ and thus, the discrete statistic $|\hat{\mathcal{B}}_k| - |\hat{\mathcal{B}}_{k+1}|$ follows the distribution $2 \cdot B(n, 1/2) - n$.

To compare the suitability of the continuous and the discrete statistics to carry steganalysis-resistant watermarks we run the steganalytic algorithm on the test set of clean models with the real values substituted by random samples following the theoretical distributions of these statistics. The results are shown in Fig. 13. The x-axis indexes the models of the test set, while the y-axis shows the value of the angle θ in Eq. 6. We notice that in the case of the Bates distribution, corresponding to the continuous statistic, the values of θ have low variance and all of them are comfortably below the $\theta_T = 2.8°$ threshold we used for watermark detection. In contrast, the angles for the discrete statistic have higher variance, and reach values close to $4°$, meaning that the watermark detection test will be less accurate in classifying these models as clean.

Notice that the above test evaluates the suitability of a statistic as steganalysis-resistant watermark carrier, it does not evaluate specific watermarking algorithms. Indeed, as clean only models are used, the test could have been run

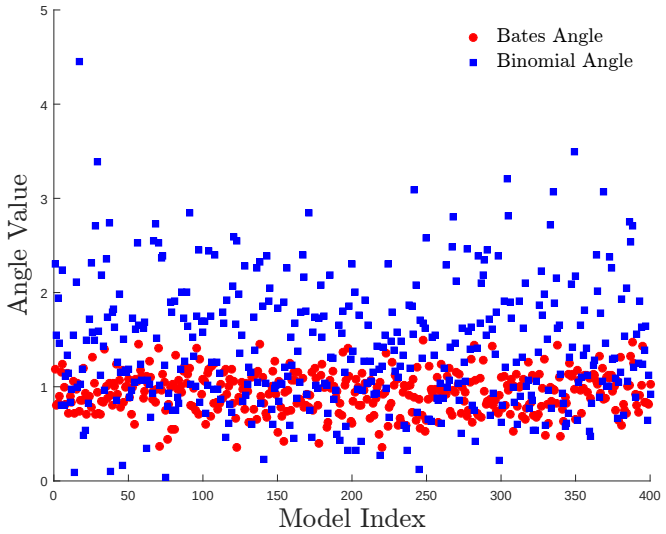


Fig. 13. The angles θ in Eq. 6 for the clean models with data sampled from the theoretical distributions. The angle values for the continuous case (Cho's et. al algorithm) are shown in red and the angles for the discrete case (proposed algorithm) are shown in blue.

at the design stage, before the development of the watermarking algorithm. However, in our case we run this test in retrospect, as a confirmation that the choice of the discrete statistic in which we based the proposed watermarking algorithm was an appropriate one.

5 Conclusion

As demonstrated by the experimental results, the developed steganalysis is able to detect the presence of watermarks by Cho et al.'s algorithms and their variants with an accuracy of up to 99%. While the proposed steganalytic algorithm was specifically designed to target Cho et al.'s two watermarking algorithms [14], the main idea could possibly be applied on several other algorithms that embed watermarks by altering a specific statistic of a histogram of the vertex set of the model.

The proposed watermarking algorithm is based on a discrete statistic of the histogram of radial coordinates, the difference in the height of adjacent bins. The experimental results demonstrate that it outperforms Yang et al.'s algorithms [7] and Cho et al.'s [14] in terms of robustness against malicious watermark removal attacks and against steganalytic attacks, while at the same time it also offers some improvement in terms of embedding distortion.

In the future, we will work to develop more advanced steganalytic techniques for detecting the presence of watermark messages embedded in 3D models, and parallelly, develop 3D watermarking/steganographic algorithms that not only have better anti-steganalytic behavior, but also offer improved robustness/distortion trade-offs.

References

[1] E. Praun, H. Hoppe, and A. Finkelstein, "Robust mesh watermarking," in *SIGGRAPH*, 1999, pp. 49–56.

[2] H.-Y. Lin, H.-Y. Liao, C.-S. Lu, and J.-C. Lin, "Fragile watermarking for authenticating 3-D polygonal meshes," *IEEE Trans. on Multimedia*, vol. 7, no. 6, pp. 997–1006, 2005.

[3] S. Zafeiriou, A. Tefas, and I. Pitas, "Blind robust watermarking schemes for copyright protection of 3D mesh objects," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 11, no. 5, pp. 596–607, 2005.

[4] Y. Yang and I. Ivrissimtzis, "Polygonal mesh watermarking using Laplacian coordinates," *Computer Graphics Forum (Proc. Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP 2010)*, vol. 29, no. 5, pp. 1585–1593, 2010.

[5] M. Luo and A. G. Bors, "Surface-preserving robust watermarking of 3-D shapes," *IEEE Trans. on Image Processing*, vol. 20, no. 10, pp. 2813–2826, 2011.

[6] K. Wang, G. Lavoué, F. Denis, and A. Baskurt, "A comprehensive survey on three-dimensional mesh watermarking," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1513–1527, 2008.

[7] Y. Yang, R. Pintus, H. Rushmeier, and I. Ivrissimtzis, "A steganalytic algorithm for 3D polygonal meshes," in *IEEE International Conference on Image Processing (ICIP) 2014*, Oct 2014, pp. 4782–4786.

[8] Y. Yang and I. Ivrissimtzis, "Mesh discriminative features for 3D steganalysis," *ACM Trans. on Multimedia Computing, Comm & Applications*, vol. 10, no. 3, pp. 27:1–27:13, 2014.

[9] A. D. Ker, "Improved detection of LSB steganography in grayscale images," in *Information Hiding*. Springer, 2005, pp. 97–115.

[10] W.-N. Lie and G.-S. Lin, "A feature-based classification technique for blind image steganalysis," *IEEE Trans. on Multimedia*, vol. 7, no. 6, pp. 1007–1020, 2005.

[11] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE TIFS*, vol. 1, no. 1, pp. 111–119, 2006.

[12] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 1, pp. 31–45, 2007.

[13] V. Holub and J. Fridrich, "Low-complexity features for jpeg steganalysis using undecimated dct," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 2, pp. 219–228, Feb 2015.

[14] J.-W. Cho, R. Prost, and H.-Y. Jung, "An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms," *IEEE Trans. on Signal Processing*, vol. 55, no. 1, pp. 142–155, 2007.

[15] B. L. Yeo and M. M. Yeung, "Watermarking 3D objects for verification," *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 36–45, 1999.

[16] Z. Yu, H. S. Ip, and L. F. Kwok, "A robust watermarking scheme for 3D triangular mesh models," *Pattern Recognition*, vol. 36, no. 11, pp. 2603–2614, 2003.

[17] A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," *IEEE Trans. on Image Processing*, vol. 15, no. 3, pp. 687–701, 2006.

[18] Z. Karni and C. Gotsman, "Spectral compression of mesh geometry," in *SIGGRAPH*, 2000, pp. 279–286.

[19] R. Ohbuchi, A. Mukaiyama, and S. Takahashi, "A frequency-domain approach to watermarking 3D shapes," *Computer Graphics Forum*, vol. 21, no. 3, pp. 373–382, 2002.

[20] S. Kanai, H. Date, and T. Kishinami, "Digital watermarking for 3D polygons using multiresolution wavelet decomposition," in *Proc. Int. Workshop on Geometric Modeling*, vol. 5, 1998, pp. 296–307.

[21] F. Uccheddu, M. Corsini, and M. Barni, "Wavelet-based blind watermarking of 3D models," in *Proc. Workshop on Multimedia and security*, 2004, pp. 143–154.

[22] F. Cayre and B. Macq, "Data hiding on 3-D triangle meshes," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 939–949, 2003.

[23] C.-M. Wang and Y.-M. Cheng, "An efficient information hiding algorithm for polygon models," *Computer Graphics Forum*, vol. 24, no. 3, pp. 591–600, 2005.

[24] Y.-M. Cheng and C.-M. Wang, "An adaptive steganographic algorithm for 3d polygonal meshes," *The Visual Computer*, vol. 23, no. 9, pp. 721–732, 2007.

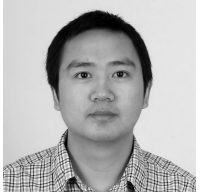
[25] M.-W. Chao, C.-H. Lin, C.-W. Yu, and T.-Y. Lee, "A high capacity 3D steganography algorithm," *IEEE Trans. on Visualization and Computer Graphics*, vol. 15, no. 2, pp. 274–284, 2009.

[26] Y. Yang, N. Peyerimhoff, and I. Ivrissimtzis, "Linear correlations between spatial and normal noise in triangle meshes," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 1, pp. 45–55, 2013.

[27] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color, and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, 2001.

[28] A. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 441–444, 2005.

[29] H. Farid, "Detecting hidden messages using higher-order statistical models," in *International Conference on Image Processing (ICIP)*, vol. 2, 2002, pp. 905–908.
[30] G. Xuan, Y. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen, "Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions," in *Proc. Information Hiding Workshop*. Springer, 2005, pp. 262–277.
[31] G. Coatrieux, W. Pan, N. Cuppens-Boulahia, F. Cuppens, and C. Roux, "Reversible watermarking based on invariant image classification and dynamic histogram shifting," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 111–120, Jan 2013.
[32] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–362, 2006.
[33] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, 2000.
[34] E. Choi and C. Lee, "Feature extraction based on the Bhattacharyya distance," *Pattern Recognition*, vol. 36, no. 8, pp. 1703–1709, 2003.
[35] B. Li, J. He, J. Huang, and Y. Q. Shi, "A survey on image steganography and steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, 2011.
[36] M. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis for the analysis of data," *Biometrika*, vol. 55, no. 1, pp. 1–17, 1968.
[37] D. Zwillinger, *CRC standard mathematical tables and formulae*. CRC Press LLC, 2003.
[38] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," in *SIGGRAPH*, 2009, pp. 73:1–73:12.
[39] G. Lavoué, "A multiscale metric for 3D mesh visual quality assessment," in *Computer Graphics Forum*, vol. 30, no. 5. Wiley Online Library, 2011, pp. 1427–1437.
[40] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: measuring error on simplified surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167–174, 1998.
[41] K. Wang, G. Lavoué, F. Denis, A. Baskurt, and X. He, "A benchmark for 3D mesh watermarking," in *Shape Modeling International*. IEEE, 2010, pp. 231–235.
[42] G. Taubin, "Geometric signal processing on polygonal meshes," *Eurographics STAR*, pp. 81–96, 2000.



Ying Yang joined the Computer Graphics Group at Yale as a Postdoctoral Associate in March 2013. Yang received the B.E. degree in Information Security in 2006 and M.E. degree in Computer Science and Technology in 2009 from the School of Computer and Communication, Hunan University, China. From September 2009 to February 2013, he pursued his PhD at the School of Engineering and Computing Sciences, Durham University UK. His research interests include digital watermarking/steganography, steganalysis, document and 3D shape analysis, and applications of computer graphics in cultural heritage.

Ruggero Pintus received the master's degree and the Ph.D. in Electronic Engineering at the University of Cagliari (Italy), with a dissertation on Computer Vision algorithms applied to Scanning Electron Microscope. He worked at the Hewlett-Packard Laboratories (Palo Alto, California) on photometric stereo techniques applied to conventional flatbed scanners. Since 2007, he has been part of the Visual Computing group of CRS4. The primary research activity was the development of algorithms for acquisition, out-of-core processing, time-critical rendering and 3D printing of massive models, mostly applied to large scale Cultural Heritage datasets. In 2013 he joined the Computer Graphics Group at Yale. His research focused on geometry and color/spectral acquisition and processing (3D Scanning, Multi-Spectral Imaging and Reflectance Transformation Imaging), and the development of algorithms for document layout analysis applied to old handwritten manuscripts.

Holly Rushmeier is a professor of Computer Science at Yale University. She received the BS, MS and PhD degrees in Mechanical Engineering from Cornell University in 1977, 1986 and 1988 respectively. Between receiving the PhD and arriving at Yale she held positions at Georgia Tech, NIST and IBM Watson research. Her current research interests include acquiring and modeling material appearance, applications of human perception to realistic rendering and applications of computer graphics in cultural heritage. She is a fellow of the Eurographics Association, an ACM Distinguished Engineer and the recipient of the 2013 ACM SIGGRAPH Computer Graphics Achievement Award.

Ioannis Ivrissimtzis is a lecturer at the School of Engineering and Computing Sciences at Durham University, UK. His research interests include subdivision surfaces, surface reconstruction from unorganised point sets and machine learning applications in computer graphics problems.